



# Magda Distributed Data Manager Status

**Torre Wenaus**

**BNL**

ATLAS Data Challenge Workshop

Feb 1, 2002

CERN

# Magda



- ❄ MAnager for Grid-based DAta
- ❄ Designed for 'managed production' *and* 'chaotic end-user' usage
- ❄ Designed for rapid development of components to support users quickly, with components later replaced by Grid Toolkit elements
  - ❑ Deploy as an evolving production tool and as a testing ground for Grid Toolkit components
- ❄ Adopted by ATLAS for 2002 ATLAS Data Challenges
- ❄ Developers - T. Wenaus and soon W. Deng (pdoc) and new hire

Info: <http://www.usatlas.bnl.gov/magda/info>

The system: <http://www.usatlas.bnl.gov/magda/dyShowMain.pl>

# Architecture & Schema



## ❄ MySQL database at the core of the system

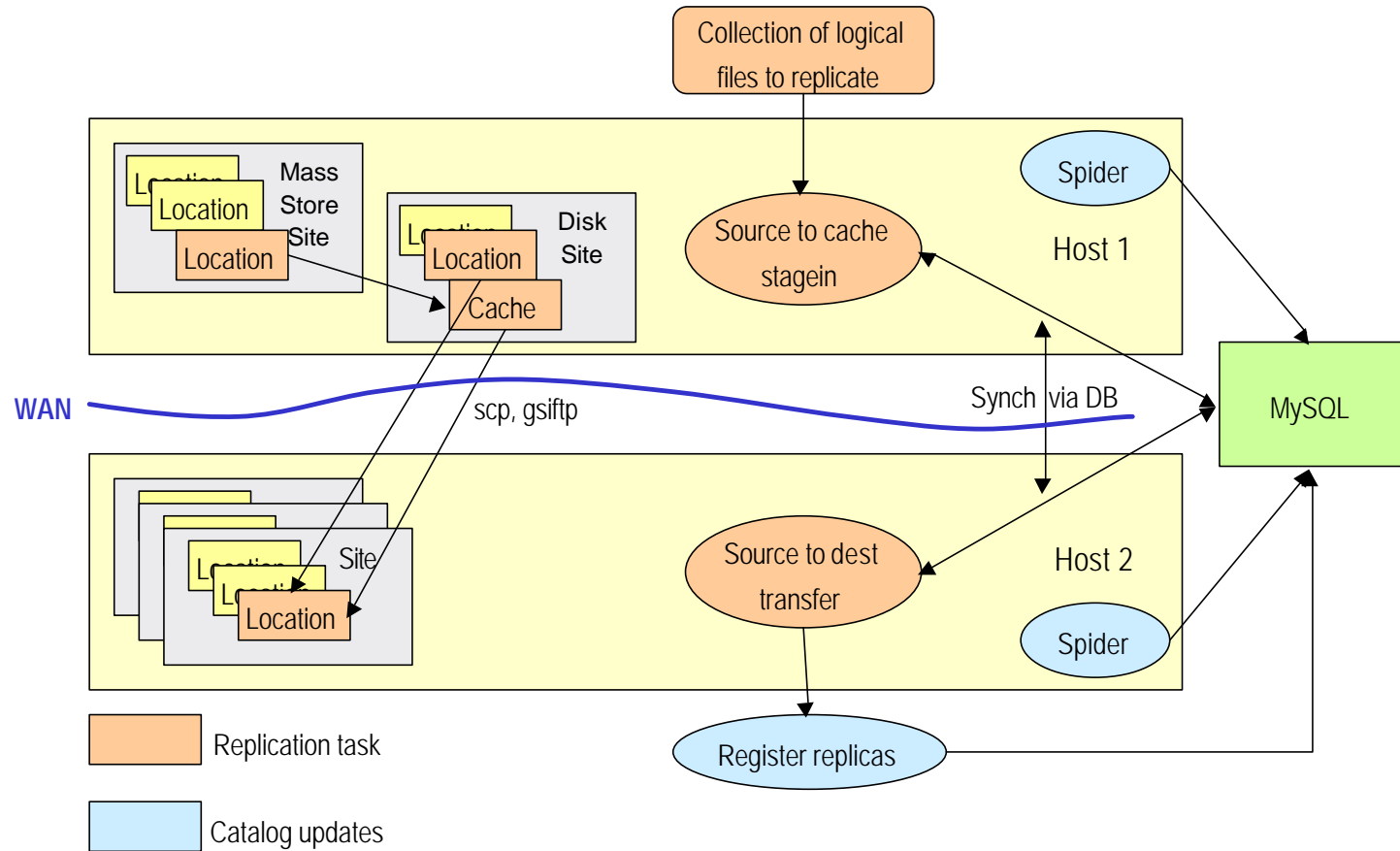
- ❑ DB interaction via perl, C++, java, cgi (perl) scripts
- ❑ C++ and Java APIs autogenerated off the MySQL DB schema

## ❄ User interaction via web interface and command line

## ❄ Principal components:

- ❑ **File catalog** covering any file types
- ❑ **Data repositories** organized into **sites**, each with its **locations**
- ❑ Computers with repository access: a **host** can access a set of **sites**
- ❑ Logical files can optionally be organized into **collections**
- ❑ **Replication** operations organized into **tasks**

# Magda Architecture



# Files and Collections



## ❄ Files & replicas

- ❑ Logical name is arbitrary string, usually but not necessarily the filename
  - ⌘ In some cases with partial path (eg. for code, path in CVS repository)
- ❑ Logical name plus virtual organization (=atlas.org) defines unique logical file
- ❑ File instances include a replica number
  - ⌘ Zero for the **master** instance; N=locationID for other instances
- ❑ Notion of **master instance** is essential for cases where replication must be done off of a specific (trusted or assured current) instance
  - ⌘ Not currently supported by Globus replica catalog

## ❄ Several types of file collections

- ❑ Logical collections: arbitrary user-defined set of logical files
- ❑ Location collections: all files at a given location
- ❑ Key collections: files associated with a key or SQL query

# Distributed Catalog



- ❄ Catalog of ATLAS data at CERN, BNL (also LBNL, ANL, BU, UTA)
  - ❑ Supported data stores: CERN Castor, CERN stage, BNL HPSS (rftp service), disk, code repositories, web areas
  - ❑ Current content: physics TDR data, test beam data, ntuples, ...
    - ⌘ About 200k files currently cataloged representing >6TB data
    - ⌘ Has run without problems with ~1.5M files cataloged
- ❄ 'Spider' crawls data stores to populate and validate catalogs
  - ❑ Catalog entries can also be added or modified directly
- ❄ Single MySQL DB serves entire system in present implementation
  - ❑ 'MySQL accelerator' provides good catalog loading performance over WAN; 2k files in <1sec. Sends bunched actions and initiates remotely with cgi
- ❄ Globus replica catalog 'loader' written for evaluation; not used yet

Magda distributed data manager prototype for ATLAS - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search History Print Print Preview Save Save As Open Recent

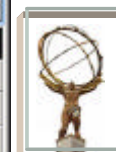
Address http://atlasw1.phy.bnl.gov/magda/dyShowMain.pl Go

Google save window to clipbo Search Web Search Site PageRank

### Primary locations containing master instances of files:

Location Name	Site:Location	Store	Type	Persist	Files	Latest file
CERN ATLAS AFS data area	<a href="cermatlasafs:/afs/cern.ch/atlas/offline/data">cermatlasafs:/afs/cern.ch/atlas/offline/data</a>	afs	data	prime	<a href="#">34</a>	2001-12-19 17:33:21
ATLAS web area	<a href="cermatlasafs:/afs/cern.ch/atlas/www">cermatlasafs:/afs/cern.ch/atlas/www</a>	afs	web	prime	<a href="#">86250</a>	2002-01-07 01:00:04
DC0 data	<a href="cerncastor:/castor/cern.ch/atlas/dc0">cerncastor:/castor/cern.ch/atlas/dc0</a>	castor	data	prime	<a href="#">1942</a>	2002-01-07 02:30:00
emebh6 testbeam	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/emebh6">cerncastor:/castor/cern.ch/atlas/testbeam/emebh6</a>	castor	data	prime	<a href="#">4095</a>	2001-10-23 15:36:00
emech6 testbeam	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/emech6">cerncastor:/castor/cern.ch/atlas/testbeam/emech6</a>	castor	data	prime	<a href="#">5682</a>	2001-10-03 17:53:00
HEC testbeam	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/hectestb">cerncastor:/castor/cern.ch/atlas/testbeam/hectestb</a>	castor	data	prime	<a href="#">888</a>	2001-12-19 07:31:00
Pixel testbeam	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/pixel">cerncastor:/castor/cern.ch/atlas/testbeam/pixel</a>	castor	data	prime	<a href="#">185</a>	2001-10-17 14:42:00
SCT testbeam	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/sct">cerncastor:/castor/cern.ch/atlas/testbeam/sct</a>	castor	data	prime	<a href="#">4389</a>	2001-12-19 09:46:00
Muon testbeam data	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/tbmMuon">cerncastor:/castor/cern.ch/atlas/testbeam/tbmMuon</a>	castor	data	prime	<a href="#">522</a>	2001-12-20 14:37:00
blecal testbeam	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/blecal">cerncastor:/castor/cern.ch/atlas/testbeam/blecal</a>	castor	data	prime	<a href="#">4496</a>	2001-12-20 10:02:00
TRT testbeam	<a href="cerncastor:/castor/cern.ch/atlas/testbeam/trttb">cerncastor:/castor/cern.ch/atlas/testbeam/trttb</a>	castor	data	prime	<a href="#">203</a>	2001-11-03 18:38:00
LAr test beam emebh6 in HPSS	<a href="cernhpss:/hpss/cern.ch/atlas/testbeam/emebh6">cernhpss:/hpss/cern.ch/atlas/testbeam/emebh6</a>	cernhpss	data	prime	<a href="#">3398</a>	2001-05-24 00:00:00
CERN stage pool: inde	<a href="cernstage:/atlas_inde">cernstage:/atlas_inde</a>	stage	data	prime	<a href="#">1110</a>	2001-12-02 19:14:36
CERN stage pool: larg	<a href="cernstage:/atlas_larg">cernstage:/atlas_larg</a>	stage	data	prime	<a href="#">321</a>	2001-08-24 16:16:15
CERN stage pool: muon	<a href="cernstage:/atlas_muon">cernstage:/atlas_muon</a>	stage	data	prime	<a href="#">35</a>	2001-07-06 17:08:09
CERN stage pool: shw	<a href="cernstage:/atlas_shw">cernstage:/atlas_shw</a>	stage	data	prime	<a href="#">1775</a>	2001-10-16

Internet



Hosts - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Reload Home Search History Print Page Setup

Address http://atlassw1.phy.bnl.gov/magda/cyShowHosts.pl Go

Google save window to clipboard Search Web Search Site PageRank

save window to clipboard save window to clipboard to clipboard

Magda info [Contact](#)

## Hosts

[PPDG at BNL](#) [BNL PAS Group](#)

Logged in as **guest** Generated Sun Jan 6 23:54:18 2002

[Main](#) - [Sites](#) - [Hosts](#) - [Collections](#) - [Add](#) - [Replication](#) - [Recent](#) - [Summary info](#) - [Web log](#) - [Admin](#)

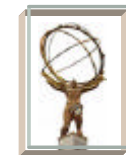
Active in last ● <24 hours, ● <1 week, ● <1 month, ● >1 month

Host	Title	Sites	Last active	VO
atlas.cern.ch	CERN ATLAS	cernatlas,cernatlasafs,cermstage,cermcastor,cermhps	<span style="color: green;">●</span> 2002-01-07 01:20:30	atlas.org
atlas.hep.anl.gov	ANL grid node	anl	<span style="color: green;">●</span> 2002-01-06 01:20:04	atlas.org
atlas.uits.iupui.edu	IU Prototype Tier 2 Center	indianaui	<span style="color: red;">●</span> 2001-11-21 17:20:51	atlas.org
atlas00.usatlas.bnl.gov	US ATLAS ssh/scp gateway	usatlas,usatlasafs,usatlasrftp	<span style="color: green;">●</span> 2002-01-06 02:11:11	atlas.org
atlasgrid.bu.edu	Boston U grid node	bostonu	<span style="color: grey;">●</span> 2001-12-15 01:20:05	atlas.org
atlassw1.phy.bnl.gov	US ATLAS Software Server	atlassw1.phy.bnl.gov	<span style="color: green;">●</span> 2002-01-06 17:01:01	atlas.org
atlastw.phy.bnl.gov	Wenus laptop	atlastw.phy.bnl.gov	<span style="color: grey;">●</span> 2001-12-23 10:24:58	atlas.org
grid.usatlas.bnl.gov	US ATLAS grid node	usatlas,usatlasafs,usatlasrftp	<span style="color: grey;">●</span> 2001-12-10 22:49:31	atlas.org
pdsfgrid.nersc.gov	PDSF grid host	lbl	<span style="color: grey;">●</span> 2001-12-10 22:49:31	atlas.org
usatlas.bnl.gov	US ATLAS Tier 1	usatlas,usatlasafs,usatlasrftp	<span style="color: green;">●</span> 2002-01-06 18:21:02	atlas.org

Internet







Add or Edit Sites, Locations, Hosts, Collections - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Search History Home

Address http://atlasw1.phy.bnl.gov/magda/cj/Add.pl

Google save window to clipboard Search Web Search Site PageRank

### Add Site

A site is a computing facility (or subset thereof) that has a number of data storage locations associated with it. eg. university X might have one site for local disks and another site for the local HPSS system.

Descriptive name of the site:  eg. 'Indiana University'

Short ID name:  eg. 'indianaui'

Virtual organization (VO):  atlas.org

Will the data at this site be public or private?

Site owner (if private):  Username

### Add Location

A location is a physical place where data is kept. eg a directory (or directory tree) on disk.

Descriptive name of the location:  eg. '2001 blecal data'

Site at which the location resides:

Location (generally, the path):  eg. /atlas/data/blecal

Virtual organization (VO):  atlas.org

Location type:

Store type:

Data type:

Include all files down the location's directory tree?

Will the data at this location be public or private?

Location owner (if private):  Username

Filter:  (optional)  
A regular expression. If filename matches, it is put in the catalog.  
eg. for 'code' type areas a filter of '!/CVS/|/configure' is used to exclude CVS directories and configure files from the catalog.

Exclude:  (optional)  
A comma separated list of subdirectories (relative to the location top level) to be excluded. Applies only to locations which include files down through

Internet

## Other Metadata



- ❄ Simple user-defined metadata support: 'keys' (strings) can be assigned to logical files
- ❄ Will integrate with external application metadata catalogs for 'metadata about the data' (eg. physics generator, run type, ...)
  - ❑ In ATLAS, a MySQL/phpMyAdmin based tool being developed by Grenoble for DC1
- ❄ Parenthetically...
  - ❑ New Magda derivative begun: **Hemp**, Hybrid Event store Metadata Prototype, for the RDBMS part of a ROOT/RDBMS event store
    - ⌘ Close ties to data signature work ('history info') as well as file management

# File Replication



- ❄ Replication operations organized as user-defined tasks specifying source collection and host, transfer tool, pull/push, destination host and location, and intermediate caches
- ❄ User-specified logical file collections are replicated
  - ❑ e.g. a set of files with a particular physics channel key
- ❄ Designed to support multiple file transfer tools, user-selectable, which are useful in different contexts (eg. scp for transfers 'outside the grid')
- ❄ In use between CERN, BNL, and among US ATLAS testbed sites
  - ❑ CERN stage, Castor, HPSS  $\Rightarrow$  cache  $\Rightarrow$  **scp**  $\Rightarrow$  cache  $\Rightarrow$  BNL HPSS
  - ❑ BNL HPSS or disk  $\Leftrightarrow$  cache  $\Leftrightarrow$  **gsiftp**  $\Leftrightarrow$  testbed disk
- ❄ ~270GB replicated to date
- ❄ GDMP integration just underway

# Replication Steps



- ❄ Replication steps for each logical file, coordinated via state info in DB:
  - ❑ Mark as 'processing' in DB collection
  - ❑ Find the least-cost replica instance accessible at source host (ie. disk instance preferred over MSS); stage into cache if necessary
  - ❑ On stage complete, mark as available for transfer
  - ❑ Independent transfer script (running on source or destination side) transfers files as they become available, and marks as available on destination side
  - ❑ If final destination is MSS, transferred files are deposited in a cache, and an independent destination-side script archives them
- ❄ Caches have 'maximum size' to throttle to available space
- ❄ If any stage breaks, others wait until file flow resumes and then proceed
- ❄ File validation is by checking file size
  - ❑ Failed transfers are re-queued

Data Replication Tasks - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Reload Search History Print Mail Print Preview

Address http://atlasw1.phy.bnl.gov/magda/dyshowTasks.pl Go

Google save window to clipboard Search Web Search Site Firefox

save window save window to clipboard

Magda info [Contact](#)

## Data Replication Tasks

EPDG at BNL BNL PAS Group

Logged in as **guest** Generated Sun Jan 6 23:39:11 2002

[Main](#) - [Sites](#) - [Hosts](#) - [Collections](#) - [Add](#) - [Replication](#) - [Recent](#) - [Summary info](#) - [Web log](#) - [Admin](#)

[Add/edit replication tasks](#)

Task Name	Type	Active	Last (de)activation	VO
<b>CERN HPSS to BNL rftp</b>	replicate	yes	2002-01-03 05:34:22	atlas.org
<p><b>Source:</b> Collection <a href="#">Files to BNL</a> via cernhpss Done: 544 Todo: 866  Most recent: <a href="#">rd13_run209361.dat</a> at 2002/01/06 18:48:25 size=3564000 complete  Source host: atlas.cern.ch  Source cache: <a href="#">Cache collection cernatlasafs:/afs/cern.ch/atlas/maxidisk/d94/wenaus/cache</a>  Most recent: <a href="#">rd13_run209352.dat</a> at 2002/01/06 03:49:54 complete</p> <p><b>Transfer operation:</b> scp from source side  Destination host: atlas00.usatlas.bnl.gov  Destination cache: <a href="#">Cache collection usatlas:/usatlas/workarea/wenaus/cache/cern</a>  Most recent: <a href="#">rd13_run209662.dat</a> at 2002/01/06 01:46:23 unknown</p> <p><b>Destination:</b> Location <a href="#">usatlasrftp:/home/wenaus/lardata</a> via rftp  Most recent: <a href="#">rd13_run209361.dat</a> at 2002/01/06 18:48 size=3564000</p>				
<b>BNL rftp to ANL</b>	replicate	no	2001-12-26 08:19:22	atlas.org
<p><b>Source:</b> Collection <a href="#">Files to ANL</a> via rftp  Source host: grid.usatlas.bnl.gov  Source cache: <a href="#">Cache collection usatlas:/usatlas/workarea/wenaus/cache/anl</a></p> <p><b>Transfer operation:</b> gsiftp from source side  Destination host: atlas.hep.anl.gov</p> <p><b>Destination:</b> Location <a href="#">anl:/usatlas/projects/wenaus/replicas</a></p>				
<b>BNL rftp to BU</b>	replicate	no	2001-12-26 08:19:27	atlas.org
<p><b>Source:</b> Collection <a href="#">Files to BU</a> via rftp  Source host: grid.usatlas.bnl.gov  Source cache: <a href="#">Cache collection usatlas:/usatlas/workarea/wenaus/cache/bu</a></p> <p><b>Transfer operation:</b> gsiftp from source side  Destination host: atlasgrid.bu.edu</p> <p><b>Destination:</b> Location <a href="#">bostonu:/atlasgrid/wenaus/replicas</a></p>				

Internet



# GDMP & Magda



- ❄ Integration as a data mover underway
- ❄ Characteristics of present implementation limit scope of its application in Magda
  - ❑ One root disk directory per site
  - ❑ Subscription updates bring in all new data for a site
  - ❑ File collections not used
  - ❑ LFN fixed as 'dir/filename' (RC constraint)
  - ❑ Doesn't catalog or directly manage files in MSS
  - ❑ Write access to tmp, etc disk areas required for all GDMP users
  - ❑ System state info (in files) only available locally
- ❄ Will try it initially for managed-production transfers between large centers

# Command Line Tools



## ❑ magda\_findfile

- ⌘ Search catalog for logical files and their instances

## ❑ magda\_getfile

- ⌘ Retrieve file via catalog lookup and (as necessary) staging from MSS or (still to come) remote replication into disk cache
- ⌘ Creates local soft link to disk instance, or a local copy
- ⌘ Usage count maintained in catalog to manage deletion

## ❑ magda\_releasefile

- ⌘ Removes local soft link, decrements usage count in catalog, deletes instance (optionally) if usage count goes to zero

## ❑ magda\_putfile

- ⌘ Archive files (eg. in Castor or HPSS) and register them in catalog
- ⌘ Or, just register them (--registeronly option)

# Near Term Activity



- ❄ Application in DC0 (deployed)
  - ❑ File management in production; replication to BNL; CERN, BNL data access
- ❄ Interface with Grenoble application metadata catalog
  - ❑ Need Grenoble API in order eg. to build file collections for replication based on application metadata selections
- ❄ GDMP integration - to be ready for DC1
  - ❑ Discussing GDMP feature set, integration issues with Heinz Stockinger next week
- ❄ Application in DC1 (beginning mid April?)
  - ❑ As DC0, but add replication and end-user data access at testbed sites
- ❄ Interface with hybrid ROOT/RDBMS event store
- ❄ Athena (ATLAS offline framework) integration; further grid integration